GenCore version 4.5

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:26:02 ; Search time 29.37 Seconds
(without alignments)
31.887 Million cell updates/sec

Title: US-09-331-631A-33
Perfect score: 77
Sequence: 1 CXXCXXXCXXXXXXXXXXXXCXXXCXXXC 29

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	77	100.0	43	1	MT2_DROME	P11956 drosophila
2	68	88.3	55	1	M84C_DROME	Q01644 drosophila
3	68	88.3	63	1	M84A_DROME	Q01642 drosophila
4	68	88.3	68	1	M84D_DROME	Q01645 drosophila
5	68	88.3	74	1	M84B_DROME	Q01643 drosophila
6	68	88.3	74	1	WDNM_MOUSE	Q62477 mus musculu
7	68	88.3	74	1	WDNM_RAT	P14730 rattus norv
8	68	88.3	95	1	EP4_RABIT	Q28631 oryctolagus
9	68	88.3	96	1	RSI1_LYCES	P47926 lycopersico
10	68	88.3	98	1	GAS1_ARATH	P46689 arabidopsis
11	68	88.3	99	1	GAS2_ARATH	P46688 arabidopsis
12	68	88.3	99	1	GAS3_ARATH	P46687 arabidopsis
13	68	88.3	106	1	GAS4_ARATH	P46690 arabidopsis
14	68	88.3	110	1	IBP_CARCR	P00993 caretta car
15	68	88.3	112	1	GST1_LYCES	P27057 lycopersico
16	68	88.3	124	1	EP4_CANFA	Q28894 canis famil
17	68	88.3	125	1	EP4_HUMAN	Q14508 homo sapien
18	68	88.3	131	1	VWF_BOVIN	P80012 bos taurus
19	68	88.3	157	1	VE5_RHPV1	P24834 rhesus papi
20	68	88.3	169	1	KRUA_HUMAN	P26371 homo sapien
21	68	88.3	182	1	KRUC_SHEEP	P26372 ovis aries
22	68	88.3	210	1	YQEH_ECOLI	Q46941 escherichia
23	68	88.3	676	1	KALM_CHICK	P33005 gallus gall
24	68	88.3	680	1	KALM_HUMAN	P23352 homo sapien
25	68	88.3	703	1	EGFR_CHICK	P13387 gallus gall
26	68	88.3	1166	1	XMRK_XIPMA	P13388 xiphophorus
27	68	88.3	1210	1	EGFR_HUMAN	P00533 homo sapien
28	68	88.3	1210	1	EGFR_MOUSE	Q01279 mus musculu
29	68	88.3	1308	1	ERB4_HUMAN	Q15303 homo sapien
30	68	88.3	1339	1	ERB3_RAT	Q62799 rattus norv
31	68	88.3	1342	1	ERB3_HUMAN	P21860 homo sapien
32	68	88.3	1700	1	BAR3_CHITE	Q03376 chironomus
33	68	88.3	2476	1	ZAN_PIG	Q28983 sus scrofa
34	68	88.3	2813	1	VWF_CANFA	Q28295 canis famil
35	68	88.3	2813	1	VWF_HUMAN	P04275 homo sapien
36	67	87.0	71	1	MT21_MYTED	P80251 mytilus edu
37	67	87.0	71	1	MT22_MYTED	P80252 mytilus edu
38	67	87.0	71	1	MT23_MYTED	P80253 mytilus edu
39	67	87.0	642	1	YB65_SCHPO	Q09746 schizosacch
40	67	87.0	704	1	KPC1_CAEEL	P34722 caenorhabdi
41	67	87.0	912	1	KPCM_HUMAN	Q15139 homo sapien
42	67	87.0	988	1	PCK1_SCHPO	P36582 schizosacch
43	67	87.0	1096	1	KPC1_ASPNG	Q00078 aspergillus
44	67	87.0	1097	1	KPC1_CANAL	P43057 candida alb
45	67	87.0	1142	1	VGLM_PHV	P27315 prospect hi

ALIGNMENTS

RESULT 1
MT2_DROME
ID MT2_DROME STANDARD; PRT; 43 AA.
AC P11956; Q9VDM2;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE METALLOTHIONEIN 2 (MT-2).
GN MTNB OR MTO.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RX MEDLINE=87204190; PubMed=3106973;
RA Mokdad R., Debec A., Wegnez M.;
RT "Metallothionein genes in Drosophila melanogaster constitute a dual
RT system.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:2658-2662(1987).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 3-27.
RC STRAIN=OREGON-R;
RX MEDLINE=91012582; PubMed=1976815;
RA Silar P., Theodore L., Mokdad R., Erraiss N.-E., Cadic A., Wegnez M.;
RT "Metallothionein Mto gene of Drosophila melanogaster: structure and
RT regulation.";
RL J. Mol. Biol. 215:217-224(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

```
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: THIS PROTEIN BINDS CATIONS OF SEVERAL TRANSITION
CC       ELEMENTS. THOUGHT TO BE INVOLVED IN METAL ION HOMEOSTASIS.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED PREDOMINANTLY IN EMBRYONIC AND
CC       LARVAL STAGES.
CC   -!- INDUCTION: STRONGLY INDUCED BY CADMIUM, COPPER AND MERCURY.
CC   -!- MISCELLANEOUS: ALL CYSTEINE RESIDUES ARE ARRANGED IN C-X-C GROUPS.
CC       THESE ARE THOUGHT TO BE THE METAL-BINDING SITES IN OTHER
CC       METALLOTHIONEINS.
CC   -!- SIMILARITY: THE TWO DROSOPHILA METALLOTHIONEINS (MTO AND MTN) ARE
CC       ONLY 26% SIMILAR.
CC   -!- SIMILARITY: BELONGS TO FAMILY 5 IN METALLOTHIONEIN SUPERFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M16250; AAA28683.1; -.
DR   EMBL; X52098; CAA36318.1; -.
DR   EMBL; AE003730; AAF55768.1; -.
DR   PIR; S14706; SMFF2.
DR   FLYBASE; FBgn0002869; MtnB.
DR   INTERPRO; IPR000966; -.
DR   PFAM; PF02067; Metallothio_5; 1.
DR   PRINTS; PR00872; MTDIPTERA.
KW   Metal-binding; Chelation; Copper; Cadmium; Zinc.
SQ   SEQUENCE   43 AA;  4525 MW;  5EE0CF9171BD9A97 CRC64;


  Query Match             100.0%;  Score 77;  DB 1;  Length 43;
  Best Local Similarity    20.7%;  Pred. No. 4.2;
  Matches    6;  Conservative   23;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CXXCXXXCXXXXXXXXXXXXCXXXCXXXC 29
           |::|:::|:::::::::::::|:::|:::|
Db       3 CKGCGTNCQCSAQKCGDNCACNKDCQCVC 31


RESULT   2
M84C_DROME
ID   M84C_DROME     STANDARD;      PRT;     55 AA.
AC   Q01644; Q9VIA0;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   MALE SPECIFIC SPERM PROTEIN MST84DC.
GN   MST84DC.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE=92102953; PubMed=1684716;
RA   Kuhn R., Kuhn C., Boersch D., Glaetzer K.H., Schaefer U.,
RA   Schaefer M.;
RT   "A cluster of four genes selectively expressed in the male germ line
RT   of Drosophila melanogaster.";
RL   Mech. Dev. 35:143-151(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- TISSUE SPECIFICITY: TESTIS.
CC   -!- DEVELOPMENTAL STAGE: PRIMARY SPERMATOCYTES.
CC   -!- DOMAIN: THIS PROTEIN IS MOSTLY COMPOSED OF REPETITIVE C-G-P
CC       MOTIFS.
CC   -!- SIMILARITY: BELONGS TO THE MST(3)CGP FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X67703; CAA47939.1; -.
DR   EMBL; AE003672; AAF54025.1; -.
DR   HSSP; P01180; 1NPO.
DR   FLYBASE; FBgn0004174; Mst84Dc.
KW   Spermatogenesis; Repeat; Multigene family.
SQ   SEQUENCE   55 AA;  5225 MW;  95A12F3AEC88BD6C CRC64;


  Query Match             88.3%;  Score 68;  DB 1;  Length 55;
  Best Local Similarity   17.9%;  Pred. No. 25;
  Matches    5;  Conservative   23;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       1 CXXCXXXCXXXXXXXXXXXCXXXCXXX 28
             |::|:::|:::::::::::::|:::|:::
Db      19 CGPCGPRCGPCGSCCGPCGPCGPCCGPF 46


RESULT   3
M84A_DROME
ID   M84A_DROME     STANDARD;      PRT;    63 AA.
AC   Q01642; Q9VIA3;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   MALE SPECIFIC SPERM PROTEIN MST84DA.
GN   MST84DA.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE=92102953; PubMed=1684716;
RA   Kuhn R., Kuhn C., Boersch D., Glaetzer K.H., Schaefer U.,
RA   Schaefer M.;
RT   "A cluster of four genes selectively expressed in the male germ line
RT   of Drosophila melanogaster.";
RL   Mech. Dev. 35:143-151(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- TISSUE SPECIFICITY: TESTIS.
CC   -!- DEVELOPMENTAL STAGE: PRIMARY SPERMATOCYTES.
CC   -!- DOMAIN: THIS PROTEIN IS MOSTLY COMPOSED OF REPETITIVE C-G-P
CC       MOTIFS.
CC   -!- SIMILARITY: BELONGS TO THE MST(3)CGP FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X67703; CAA47937.1; -.
DR   EMBL; AE003672; AAF54022.1; -.
DR   HSSP; P01180; 1NPO.
DR   FLYBASE; FBgn0004172; Mst84Da.
KW   Spermatogenesis; Repeat; Multigene family.
SQ   SEQUENCE   63 AA;  5806 MW;  BF84CD74CBCEFD3F CRC64;


  Query Match             88.3%;  Score 68;  DB 1;  Length 63;
  Best Local Similarity   17.9%;  Pred. No. 27;
  Matches    5;  Conservative   23;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CXXCXXXCXXXXXXXXXXXCXXXCXXX 28
             |::|:::|:::::::::::::|:::|:::
Db      26 CGGCGPCCGGCGPCCGPCGGCGPCCGGT 53


RESULT   4
M84D_DROME
ID   M84D_DROME     STANDARD;      PRT;    68 AA.
AC   Q01645;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   MALE SPECIFIC SPERM PROTEIN MST84DD.
GN   MST84DD.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE=92102953; PubMed=1684716;
RA   Kuhn R., Kuhn C., Boersch D., Glaetzer K.H., Schaefer U.,
RA   Schaefer M.;
RT   "A cluster of four genes selectively expressed in the male germ line
RT   of Drosophila melanogaster.";
RL   Mech. Dev. 35:143-151(1991).
CC   -!- TISSUE SPECIFICITY: TESTIS.
CC   -!- DEVELOPMENTAL STAGE: PRIMARY SPERMATOCYTES.
CC   -!- DOMAIN: THIS PROTEIN IS MOSTLY COMPOSED OF REPETITIVE C-G-P
CC       MOTIFS.
CC   -!- SIMILARITY: BELONGS TO THE MST(3)CGP FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X67703; CAA47940.1; -.
DR   HSSP; P01180; 1NPO.
DR   FLYBASE; FBgn0004175; Mst84Dd.
KW   Spermatogenesis; Repeat; Multigene family.
SQ   SEQUENCE   68 AA;  6480 MW;  2F2BD26128DE3DEF CRC64;


  Query Match             88.3%;  Score 68;  DB 1;  Length 68;
  Best Local Similarity   17.9%;  Pred. No. 29;
  Matches    5;  Conservative   23;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 CXXCXXXCXXXXXXXXXXXXCXXXCXXX 28
          |::|:::|:::::::::::::|:::|:::
Db     29 CGPCGPCCGPCGPRCGPCGPCGPCCGTM 56


RESULT   5
M84B_DROME
ID   M84B_DROME     STANDARD;      PRT;    74 AA.
AC   Q01643;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   MALE SPECIFIC SPERM PROTEIN MST84DB.
GN   MST84DB.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE=92102953; PubMed=1684716;
RA   Kuhn R., Kuhn C., Boersch D., Glaetzer K.H., Schaefer U.,
RA   Schaefer M.;
RT   "A cluster of four genes selectively expressed in the male germ line
RT   of Drosophila melanogaster.";
RL   Mech. Dev. 35:143-151(1991).
CC   -!- TISSUE SPECIFICITY: TESTIS.
CC   -!- DEVELOPMENTAL STAGE: PRIMARY SPERMATOCYTES.
CC   -!- DOMAIN: THIS PROTEIN IS MOSTLY COMPOSED OF REPETITIVE C-G-P
CC       MOTIFS.
CC   -!- SIMILARITY: BELONGS TO THE MST(3)CGP FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X67703; CAA47938.1; -.
DR   HSSP; P01180; 1NPO.
DR   FLYBASE; FBgn0004173; Mst84Db.
KW   Spermatogenesis; Repeat; Multigene family.
SQ   SEQUENCE   74 AA;  6824 MW;  D0793137A7E7D1D0 CRC64;


  Query Match             88.3%;  Score 68;  DB 1;  Length 74;
  Best Local Similarity   17.9%;  Pred. No. 31;
  Matches    5;  Conservative   23;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CXXCXXXCXXXXXXXXXXXXCXXXCXXX 28
          |::|:::|:::::::::::::|:::|:::
Db     28 CGSCCSPCGSCCAPWGPCGPCGPCCGGC 55


RESULT   6
WDNM_MOUSE
ID   WDNM_MOUSE     STANDARD;      PRT;    74 AA.
AC   Q62477;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   WDNM1 PROTEIN PRECURSOR.
GN   WDNM1 OR EXPI.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=BREAST;
RX   MEDLINE=95060797; PubMed=7970700;
RA   Morrison B.W., Leder P.;
RT   "neu and ras initiate murine mammary tumors that share genetic
RT   markers generally absent in c-myc and int-2-initiated tumors.";
RL   Oncogene 9:3417-3426(1994).
CC   -!- FUNCTION: COULD HAVE PROTEINASE INHIBITING CAPACITY.
CC   -!- SIMILARITY: BELONGS TO THE WAP-TYPE 'FOUR-DISULFIDE CORE' FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X93037; CAA63605.1; -.
DR   HSSP; P19957; 2REL.
DR   MGD; MGI:107506; EXPI.
DR   INTERPRO; IPR002221; -.
DR   PFAM; PF00095; wap; 1.
DR   PRINTS; PR00003; 4DISULPHCORE.
DR   PROSITE; PS00317; 4_DISULFIDE_CORE; 1.
KW   Protease inhibitor; Signal.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25     74       WDNM1 PROTEIN.
SQ   SEQUENCE   74 AA;  7787 MW;  42BDCAD5B330B3D3 CRC64;


  Query Match             88.3%;  Score 68;  DB 1;  Length 74;
  Best Local Similarity   17.9%;  Pred. No. 31;
  Matches    5;  Conservative   23;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 XXCXXXCXXXXXXXXXXXXCXXXCXXXC 29
          ::|:::|:::::::::::::|:::|:::|
Db     42 GTCDERCTGDGSCSGNMKCCSNGCGHAC 69


RESULT   7
WDNM_RAT
ID   WDNM_RAT       STANDARD;      PRT;    74 AA.
AC   P14730;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   WDNM1 PROTEIN PRECURSOR (FRAGMENT).
GN   WDNM1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BREAST CARCINOMA;
RX   MEDLINE=88310901; PubMed=3136918;
RA   Dear T.N., Ramshaw I.A., Kefford R.F.;
RT   "Differential expression of a novel gene, WDNM1, in nonmetastatic rat
RT   mammary adenocarcinoma cells.";
RL   Cancer Res. 48:5203-5209(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=THYMUS;
RX   MEDLINE=91207400; PubMed=2018519;
RA   Dear T.N., Kefford R.F.;
RT   "The WDNM1 gene product is a novel member of the 'four-disulphide
RT   core' family of proteins.";
RL   Biochem. Biophys. Res. Commun. 176:247-254(1991).
CC   -!- FUNCTION: INVOLVED IN THE METASTATIC POTENTIAL OF ADENOCARCINOMAS
CC       IN RAT. COULD HAVE PROTEINASE INHIBITING CAPACITY.
CC   -!- TISSUE SPECIFICITY: SPLEEN, KIDNEY, BRAIN, LIVER, LUNG, AND CELL
CC       LINE RAT-2.
CC   -!- SIMILARITY: BELONGS TO THE WAP-TYPE 'FOUR-DISULFIDE CORE' FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X13309; CAA31688.1; ALT_INIT.
DR   PIR; JH0390; JH0390.
DR   HSSP; P19957; 2REL.
DR   INTERPRO; IPR002221; -.
DR   PFAM; PF00095; wap; 1.
DR   PROSITE; PS00317; 4_DISULFIDE_CORE; 1.
KW   Protease inhibitor; Signal.
FT   NON_TER       1      1
FT   SIGNAL       <1     23       POTENTIAL.
FT   CHAIN        24     74       WDNM1 PROTEIN.
SQ   SEQUENCE   74 AA;  7740 MW;  51802C70CDAF0521 CRC64;


  Query Match             88.3%;  Score 68;  DB 1;  Length 74;
  Best Local Similarity   17.9%;  Pred. No. 31;
  Matches    5;  Conservative   23;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 XXCXXXCXXXXXXXXXXXCXXXCXXXC 29
           ::|:::|:::::::::::::|:::|:::|
Db      42 GTCVELCSGDQSCPNIQKCCSNGCGHVC 69


RESULT   8
EP4_RABIT
ID   EP4_RABIT      STANDARD;      PRT;    95 AA.
AC   Q28631;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   MAJOR EPIDIDYMIS-SPECIFIC PROTEIN E4 (EPIDIDYMAL PROTEIN BE-20)
DE   (FRAGMENT).
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EPIDIDYMIS;
RA   Wang L.F.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POSSIBLE FUNCTION IN SPERM MATURATION.
CC   -!- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC   -!- TISSUE SPECIFICITY: EPIDIDYMIS.
CC   -!- SIMILARITY: BELONGS TO THE WAP-TYPE 'FOUR-DISULFIDE CORE' FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U26725; AAA66525.1; -.
DR   HSSP; P19957; 1FLE.
DR   INTERPRO; IPR002221; -.
DR   PFAM; PF00095; wap; 2.
DR   PROSITE; PS00317; 4_DISULFIDE_CORE; 2.
KW   Glycoprotein.
FT   NON_TER       1      1
FT   DOMAIN       <1     47       WAP 1.
FT   DOMAIN       48     95       WAP 2.
FT   CARBOHYD     17     17       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   95 AA;  9990 MW;  58BAF723BD74D7F3 CRC64;


  Query Match             88.3%;  Score 68;  DB 1;  Length 95;
  Best Local Similarity   17.9%;  Pred. No. 36;
  Matches    5;  Conservative   23;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 XXCXXXCXXXXXXXXXXXCXXXCXXXC 29
           ::|:::|:::::::::::::|:::|:::|
Db      16 LNCTQDCRADQDCAENLKCCRAGCSAIC 43


RESULT   9
RSI1_LYCES
ID   RSI1_LYCES     STANDARD;      PRT;    96 AA.
AC   P47926;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   RSI-1 PROTEIN PRECURSOR (TR132).
GN   RSI-1.
OS   Lycopersicon esculentum (Tomato).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC   Solanales; Solanaceae; Solanum.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. VFNT; TISSUE=ROOT;
RX   MEDLINE=94232181; PubMed=8177211;
RA   Taylor B.H., Scheuring C.F.;
RT   "A molecular marker for lateral root initiation: the RSI-1 gene of
RT   tomato (Lycopersicon esculentum Mill) is activated in early lateral
RT   root primordia.";
RL   Mol. Gen. Genet. 243:148-157(1994).
CC   -!- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC   -!- TISSUE SPECIFICITY: EXPRESSED VERY EARLY IN LATERAL ROOT
CC       DEVELOPMENT.
CC   -!- INDUCTION: BY AUXINS.
CC   -!- PTM: SIX DISULFIDE BONDS MAY BE PRESENT.
CC   -!- SIMILARITY: BELONGS TO THE GAST1 FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; L22188; AAA20129.1; -.
DR   EMBL; L22189; AAA20130.1; -.
KW   Signal.
FT   SIGNAL        1     29       POTENTIAL.
FT   CHAIN        30     96       RSI-1 PROTEIN.
SQ   SEQUENCE   96 AA;  10839 MW;  C17C826648148CE3 CRC64;


  Query Match             88.3%;  Score 68;  DB 1;  Length 96;
  Best Local Similarity   17.9%;  Pred. No. 36;
  Matches    5;  Conservative   23;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 XXCXXXCXXXXXXXXXXXCXXXCXXXC 29
           ::|:::|:::::::::::::|:::|:::|
Db      36 TDCKPRCTYRCSATSHKKPCMFFCQKCC 63


RESULT  10
GAS1_ARATH
ID   GAS1_ARATH     STANDARD;      PRT;    98 AA.
AC   P46689;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GIBBERELLIN-REGULATED PROTEIN 1 PRECURSOR.
GN   GASA1.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
```

```
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA; TISSUE=SILIQUES;
RX   MEDLINE=95244835; PubMed=7727751;
RA   Herzog M., Dorne A.-M., Grellet F.;
RT   "GASA, a gibberellin-regulated gene family from Arabidopsis thaliana
RT   related to the tomato GAST1 gene.";
RL   Plant Mol. Biol. 27:743-752(1995).
RN   [2]
RP   SEQUENCE OF 57-98 FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Raynal M., Grellet F., Laudie M., Meyer Y., Cooke R., Delseny M.;
RL   Submitted (FEB-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAY FUNCTION IN SOME HORMONAL CONTROLLED STEP OF FRUIT
CC       DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN FLOWER BUDS AND GREEN SILIQUES.
CC       LOWER LEVELS SEEN IN THE ROOTS AND ROSETTES.
CC   -!- INDUCTION: BY GIBBERELLINS.
CC   -!- PTM: SIX DISULFIDE BONDS MAY BE PRESENT.
CC   -!- SIMILARITY: BELONGS TO THE GAST1 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U11766; AAB06310.1; -.
DR   EMBL; Z29940; CAA82842.1; -.
KW   Multigene family; Signal.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24     98       GIBBERELLIN-REGULATED PROTEIN 1.
SQ   SEQUENCE   98 AA;  10718 MW;  BB37B5F0A7C6C89B CRC64;


  Query Match             88.3%;  Score 68;  DB 1;  Length 98;
  Best Local Similarity   17.9%;  Pred. No. 37;
  Matches    5;  Conservative   23;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 XXCXXXCXXXXXXXXXXXXCXXXCXXXC 29
           ::|:::|:::::::::::::|:::|:::|
Db      38 IDCGSACVARCRLSRRPRLCHRACGTCC 65


RESULT  11
GAS2_ARATH
ID   GAS2_ARATH     STANDARD;      PRT;    99 AA.
AC   P46688;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   GIBBERELLIN-REGULATED PROTEIN 2 PRECURSOR.
GN   GASA2.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA; TISSUE=SEED;
RX   MEDLINE=95244835; PubMed=7727751;
RA   Herzog M., Dorne A.-M., Grellet F.;
RT   "GASA, a gibberellin-regulated gene family from Arabidopsis thaliana
RT   related to the tomato GAST1 gene.";
RL   Plant Mol. Biol. 27:743-752(1995).
CC   -!- FUNCTION: INVOLVED IN LATE STAGES OF SEED MATURATION, OR IN EARLY
CC       STEPS OF GERMINATION.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: DRY SEEDS AND MATURATING SILIQUES.
CC   -!- PTM: SIX DISULFIDE BONDS MAY BE PRESENT.
CC   -!- SIMILARITY: BELONGS TO THE GAST1 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U11765; AAB06309.1; -.
KW   Multigene family; Signal.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19     99       GIBBERELLIN-REGULATED PROTEIN 2.
SQ   SEQUENCE   99 AA;  10531 MW;  63713C986887573B CRC64;


  Query Match             88.3%;  Score 68;  DB 1;  Length 99;
  Best Local Similarity   17.9%;  Pred. No. 37;
  Matches    5;  Conservative   23;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 XXCXXXCXXXXXXXXXXXXCXXXCXXXC 29
           ::|:::|:::::::::::::|:::|:::|
Db      39 IDCGGRCKDRCSKSSRTKLCLRACNSCC 66


RESULT  12
GAS3_ARATH
ID   GAS3_ARATH     STANDARD;      PRT;    99 AA.
AC   P46687;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   GIBBERELLIN-REGULATED PROTEIN 3 PRECURSOR.
GN   GASA3.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA; TISSUE=SEED;
RX   MEDLINE=95244835; PubMed=7727751;
RA   Herzog M., Dorne A.-M., Grellet F.;
RT   "GASA, a gibberellin-regulated gene family from Arabidopsis thaliana
RT   related to the tomato GAST1 gene.";
RL   Plant Mol. Biol. 27:743-752(1995).
CC   -!- FUNCTION: INVOLVED IN LATE STAGES OF SEED MATURATION, OR IN EARLY
CC       STEPS OF GERMINATION.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: SILIQUES AND DRY SEEDS.
CC   -!- PTM: SIX DISULFIDE BONDS MAY BE PRESENT.
CC   -!- SIMILARITY: BELONGS TO THE GAST1 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U11764; AAB06308.1; -.
KW   Multigene family; Signal.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19     99       GIBBERELLIN-REGULATED PROTEIN 3.
SQ   SEQUENCE   99 AA;  10704 MW;  84549B74B08A47FA CRC64;


  Query Match             88.3%;  Score 68;  DB 1;  Length 99;
  Best Local Similarity   17.9%;  Pred. No. 37;
```

```
  Matches    5;  Conservative   23;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 XXCXXXCXXXXXXXXXXXXCXXXCXXXC 29
           ::|:::|::::::::::::|:::|:::|
Db      39 IDCGGRCKGRCSKSSRPNLCLRACNSCC 66


RESULT  13
GAS4_ARATH
ID   GAS4_ARATH     STANDARD;      PRT;   106 AA.
AC   P46690;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   GIBBERELLIN-REGULATED PROTEIN 4 PRECURSOR.
GN   GASA4.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. C24; TISSUE=FLOWER BUDS;
RX   MEDLINE=95244835; PubMed=7727751;
RA   Herzog M., Dorne A.-M., Grellet F.;
RT   "GASA, a gibberellin-regulated gene family from Arabidopsis thaliana
RT   related to the tomato GAST1 gene.";
RL   Plant Mol. Biol. 27:743-752(1995).
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: DEVELOPING ROOTS AND FLOWER BUDS.
CC   -!- INDUCTION: BY GIBBERELLINS.
CC   -!- PTM: SIX DISULFIDE BONDS MAY BE PRESENT.
CC   -!- SIMILARITY: BELONGS TO THE GAST1 FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U15683; AAA74480.1; -.
KW   Multigene family; Signal.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26    106       GIBBERELLIN-REGULATED PROTEIN 4.
SQ   SEQUENCE   106 AA;  12065 MW;  62A3E4A820F57678 CRC64;


  Query Match             88.3%;  Score 68;  DB 1;  Length 106;
  Best Local Similarity   17.9%;  Pred. No. 39;
  Matches    5;  Conservative   23;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 XXCXXXCXXXXXXXXXXXXCXXXCXXXC 29
           ::|:::|::::::::::::|:::|:::|
Db      46 TQCPSECDRRCKKTQYHKACITFCNKCC 73


RESULT  14
IBP_CARCR
ID   IBP_CARCR      STANDARD;      PRT;   110 AA.
AC   P00993;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   CHELONIANIN (BASIC PROTEASE INHIBITOR) (RTPI).
OS   Caretta caretta (Loggerhead).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Testudines; Cryptodira; Chelonioidea; Cheloniidae; Caretta.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=EGG WHITE;
RA   Kato I., Tominaga N.;
RT   "Trypsin-subtilisin inhibitor from red sea turtle eggwhite consists
RT   of two tandem domains -- one Kunitz -- one of a new family.";
RL   Fed. Proc. 38:832-832(1979).
CC   -!- MISCELLANEOUS: THIS INHIBITOR, ISOLATED FROM EGG WHITE, CONSISTS
CC       OF TWO NONHOMOLOGOUS DOMAINS.
CC   -!- SIMILARITY: THE FIRST DOMAIN BELONGS TO THE BPTI/KUNITZ FAMILY OF
CC       INHIBITORS.
CC   -!- SIMILARITY: THE SECOND DOMAIN IS A WAP-TYPE 'FOUR-DISULFIDE CORE'
CC       DOMAIN.
CC   -!- CAUTION: AS THE PAPER ONLY INDICATES THE SPECIES AS "RED SEA
CC       TURTLE", THE SPECIES INDICATED HERE IS THEREFORE AN INFERENCE.
DR   PIR; A01224; TITTOR.
DR   HSSP; P00974; 1BRB.
DR   INTERPRO; IPR002221; -.
DR   INTERPRO; IPR002223; -.
DR   PFAM; PF00014; Kunitz_BPTI; 1.
DR   PFAM; PF00095; wap; 1.
DR   PRINTS; PR00003; 4DISULPHCORE.
DR   PRINTS; PR00759; BASICPTASE.
DR   PROSITE; PS00317; 4_DISULFIDE_CORE; 1.
DR   PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR   PROSITE; PS50279; BPTI_KUNITZ_2; 1.
KW   Serine protease inhibitor.
FT   DOMAIN        1     58       INHIBIT TRYPSIN.
FT   DOMAIN       59    110       INHIBIT SUBTILISIN.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   ACT_SITE     18     19       REACTIVE BOND (TRYPSIN).
SQ   SEQUENCE   110 AA;  11916 MW;  269436243813418E CRC64;


  Query Match             88.3%;  Score 68;  DB 1;  Length 110;
  Best Local Similarity   17.9%;  Pred. No. 40;
  Matches    5;  Conservative   23;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 XXCXXXCXXXXXXXXXXXXCXXXCXXXC 29
           ::|:::|::::::::::::|:::|:::|
Db      74 GICLHGCDSDSDCKEGQKCCFDGCGYIC 101


RESULT  15
GST1_LYCES
ID   GST1_LYCES     STANDARD;      PRT;   112 AA.
AC   P27057;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   GAST1 PROTEIN PRECURSOR.
GN   GAST1.
OS   Lycopersicon esculentum (Tomato).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC   Solanales; Solanaceae; Solanum.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. MONEYMAKER; TISSUE=SHOOT;
RX   MEDLINE=93251105; PubMed=1302047;
RA   Shi L., Gast R.T., Gopalraj M., Olszewski N.E.;
RT   "Characterization of a shoot-specific, GA3- and ABA-regulated gene
RT   from tomato.";
RL   Plant J. 2:153-159(1992).
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: ALL SHOOT ORGANS.
CC   -!- INDUCTION: BY GIBBERELLINS. INHIBITED BY ABA.
CC   -!- PTM: SIX DISULFIDE BONDS MAY BE PRESENT.
CC   -!- SIMILARITY: BELONGS TO THE GAST1 FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ----------------------------------------------------------------------
DR   EMBL; X63093; CAA44807.1; -.
DR   PIR; S22151; S22151.
KW   Signal.
FT   SIGNAL        1      ?       POTENTIAL.
FT   CHAIN         ?    112       GAST1 PROTEIN.
SQ   SEQUENCE   112 AA;  12784 MW;  D369F466DF7E1C97 CRC64;


  Query Match             88.3%;  Score 68;  DB 1;  Length 112;
  Best Local Similarity   17.9%;  Pred. No. 40;
  Matches    5;  Conservative   23;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 XXCXXXCXXXXXXXXXXXXCXXXCXXXC 29
           ::|:::|:::::::::::::|:::|:::|
Db      52 QDCQPKCTYRCSKTSYKKPCMFFCQKCC 79


Search completed: March  1, 2001, 16:26:03
Job time: 400 sec
```